**Supersize It: The Growth of Retail Chains and the Rise of the "Big Box" Retail Format***

Emek Basker
University of Missouri

Shawn D. Klimek
U.S. Bureau of the Census

Pham Hoang Van
Baylor University

Census Advisory Committee of Professional Associations Meetings
April 19-20, 2007

Abstract

Retail chains have grown and retail stores have expanded their selection of products over the last several decades. We use micro data from the Census of Retail Trade to document these facts and show that these trends are related: retail chains with more stores tend to carry more distinct product lines, and as retail chains grow, they add both stores and product lines. To explain this fact, we present a model of a retail firm in which fixed costs are associated with both types of expansion – into new geographic and product markets – resulting in a complementarity between the two types of expansion.

JEL Codes: L11, L25, L81

Keywords: Retail Chains, Big Box, Economics of Scale, General Merchandise, One Stop Shopping

Discussion Questions:

1. Are the assumptions of the model feasible?
2. What is the appropriate definition of a market?
3. Is the detailed product data in Retail Census detailed enough to study the "Big Box" phenomenon?

---

*Very preliminary; please do not cite or distribute without permission from the authors. This work is unofficial and thus has not undergone the review accorded to official Census Bureau publications. The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau. Comments welcome to: emek@missouri.edu, shawn.d.klimek@census.gov, or van_pham@baylor.edu. We thank Roger Betancourt, Jonathan Beck, Tom Davidoff, Lucia Foster, Ron Jarmin, and seminar participants at Berkeley (Haas), Missouri, and the "Advances in the Empirical Analysis of Retailing" workshop at WZ-Berlin for comments. Basker also thanks the Economic and Policy Analysis Research Center (EPARC) at the University of Missouri for research support.

# 1 Introduction

The growth of "big box" retail chains — large, box-shaped general-merchandisers selling everything from toothpaste to baby strollers to vacuum cleaners and televisions — has received attention from both the popular press and academic economists. Chains like Wal-Mart, Target, and Kmart account for an ever-increasing share of Americans' purchases; in 2004, Wal-Mart alone handled 6.5% of all U.S. retail sales (Basker, 2006). These chain retailers are also expanding in product space. Most recently, all three chains have added a full line of groceries to selected stores, which had already sold apparel, housewares, prescription drugs, toys, books, and electronics. Wal-Mart, for example, introduced its "Supercenter" store format, which includes a full line of groceries, in 1988, and by 2002 was the largest grocer in the U.S. (Basker, 2006). Target, too, operates superstores, called SuperTarget® stores, and Kmart operates both Big Kmart stores with an expanded food area (but not a full grocery store) and Kmart Super Centers, which contain a full grocery store.

In this paper, we document a strong and persistent relationship between the size, or scale, of a retail chain, as measured by the number of stores the chain operates, and its scope, as measured by the number and variety of products it sells. We use newly-developed micro data from the Census of Retail Trade to show that the number of products a store sells is increasing in the size of the retail chain to which it belongs, and that as retail chains grow in scale — adding geographic markets — they also grow in scope — adding product lines. We propose a model to account for this relationship, in which there are fixed costs associated with adding both geographic markets and product markets. Because larger retailers can spread these fixed costs more thinly, we observe a complementarity between chain scale and scope.

We focus on the general merchandise subsector, in which stores sell a wide variety of products. In this subsector the rise of retail chains has been most spectacular; single-store ("mom and pop") retailers now account for under 2% of retail dollars in this subsector, as compared with over 37% in the retail sector overall.

Our paper is related to an established, largely theoretical, literature on the role of retailers, and to a more recent empirical literature on the "big box" phenomenon. The theoretical literature has focused on competition and retailer behavior within a single market — Bliss (1988), for example, discusses competition between "specialist" shops and a "general" shop, and Bagwell, Ramey, and Spulber (1997) discuss the rise of a dominant, low-cost, low-price retail firm — but neither paper discusses the special role of the retail chain. Holmes (2001) models an increase in size of a retail store caused by improvements in inventory-management technology, for example the introduction of bar codes or radio frequency identification techniques. We argue that, by increasing the optimal scope of stores, these innovations indirectly lead the chain to expand its scale, as well.[1]

The related empirical literature concerns the growth of retail chains, and their impact on local markets. Foster, Haltiwanger, and Krizan (2006) use data from the Census of Retail Trade to study productivity dynamics in the retail sector, and find that large (and expanding) retail chains are much more efficient than "mom and pop" stores. Jarmin, Klimek, and Miranda (2005) use data from the Census Bureau's Longitudinal Business Database (LBD) to study the growth of retail chains using employment per store as a measure of store size. The average number of workers per store has more than doubled in the last 40 years (going from approximate6ly 6 to approximately 14), at the same time that the share of stores operated by chains has grown dramatically — from 20% to 35% (Jarmin, Klimek, and Miranda, 2005). This trend has been even more dramatic in the general-merchandise subsector, on which we focus in this paper. The number of county markets served by at least one single-unit general merchandise retailer declined by 28% (from 2835 to 2138) between 1977 and 2002, while the number of county markets served by a local or regional general-merchandisers declined by 86% and 80% respectively. At the same time, the number of

[1] Competition between "local" and "global" players has been discussed in other contexts, most notably media markets where competition between different types of players can change their choices of product characteristics, and, in particular, lead to homogenization of products (see, e.g., Loertscher and Muehlheusser, 2006; George and Waldfogel, 2003).

county markets served by at least one general merchandiser with a national chain increased by 25% (from 2087 to 2673). Unlike Jarmin, Klimek, and Miranda, we use the number of stores operated by a retailer, rather than the number of geographic markets served, as our measure of retailer scale. When we talk about a retailer's "geographic expansion," then, we are implicitly assuming that each store is in a unique geographic market.[2]

There is also a growing literature on the effects of big box stores on local economies, to which this paper contributes indirectly. Basker (2006) reviews the literature on the local effects of big box stores with a special focus on Wal-Mart, the largest and most visible of the large retail chains. We argue in this paper that the competitive effect of a large retail store may well depend on the size of the chain it belongs to, and that as the chain grows, its competitive impact in each local market is likely to increase.

The rest of the paper is organized as follows. Section 2 provides an overview of the general-merchandise subsector and the changes it has undergone in the last few decades, as it compares to the retail sector as a whole. Section 3 describes the Census data we use. Section 4 presents our empirical findings. Section 5 presents a model to explain these results, in which a retailer's store size (scope) and chain size (scale) are complements. Section 6 concludes.

# 2 Background on General Merchandise Stores

We use the following terminology in this paper. A **retailer** is a firm which may operate one or more retail establishments, or **stores**. A retailer operating a single store is a **single-store retailer** (or "mom and pop" retailer). A retailer operating multiple stores is a **retail chain**.

The North American Industrial Classification System (NAICS), currently used by the Census, defines "General Merchandise Stores" as stores that sell "new general merchandise

[2]In other words, we are assuming that a single retailer's stores do not "self-cannibalize." This seems like a reasonable first approximation, although Holmes (2005) argues that cost-savings associated with "economies of density" can be sufficiently large to compensate for such cannibalization in practice.

[to retail consumers] from fixed point-of-sale locations. Establishments in this subsector are unique in that they have the equipment and staff capable of retailing a large variety of goods from a single location."[3] Before the classification change that took place in 1997, the Census used the Standard Industrial Classification (SIC) system, which defined the subsector as consisting of "stores which sell a number of lines of merchandise, such as dry goods, apparel and accessories, furniture and home furnishings, small wares, hardware, and food."[4] The subsector includes department stores, discount department stores, dollar stores, general stores, variety stores and trading posts.

The sales share of chain retailers has been increasing in recent years; and it is increasing even faster for large retail chains than for small ones. The first column of Table 1 documents the rising share of retail sales accounted for by retail chains. Until the late 1970s, more than half of all consumer dollars were spent at single-store ("mom and pop") retailers; today, more than 60% of consumer dollars are spend at chain stores, double the share of 1954. The third column shows the revenue share of large chains, defined as chains with 100 or more stores, which has grown even faster than the chains' share as a whole; large chains' share of retail sales has more than tripled since the 1950s.

Chains are dramatically more important in the general merchandise retail subsector, where they have essentially taken over the entire sector. The second and fourth columns in Table 1 show that in the General Merchandise (GM) subsector, where retailers carry many varieties of goods, chains have been dominant for many decades, but single-store retailers are now virtually extinct: almost 99% of all general merchandise dollars are spent in chains, and 96% in chains with 100 or more stores (up from 34% fifty years ago).[5]

Table 2 provides some statistics on the differences between the retail sector as a whole and the general merchandise subsector from the 2002 Census of Retail Trade (CRT). General

[3] `http://www.census.gov/epcd/naics02/def/NDEF452.HTM`

[4] `http://www.census.gov/epcd/ec97sic/def/G53.TxT`

[5] Prior to 1900, chains' share of sales was below 10% even the grocery subsector, where chains grew very early (Barger, 1955).

merchandise retailers constitute only 1.3% of all retail firms, but account for 3.7% of retail stores and 14.6% of retail sales. In 2002, the average general merchandise retailer had a chain of 3.8 stores, while the average retailer had a chain of 1.3 stores. More strikingly, while 75% of retail firms operated a single store, 72% of general-merchandise stores belonged to chains with 100 or more stores.

# 3 Data

We use micro data from the Census of Retail Trade (CRT) for the years 1982, 1987, 1992, 1997, and 2002. Detailed CRT forms, which include the data we use, are mailed to every chain store, as well as to a sample of single-unit retailers with more than 10 employees.[6] The CRT allows us to identify general-merchandise stores, link all stores within a chain, and track them over time. In addition, we have information on the number, and type, of products sold by each store at each point in time, as reported by the store's manager.[7]

We limit our analysis to stores that received Census forms numbered 5301 or 5302 (in 1982-1997) or 45201 or 45202 (in 2002). These forms are designated for general merchandise stores, although a small fraction of general merchandise stores receive other forms (for example, forms intended for apparel stores or supermarkets). We also eliminate from the analysis stores that received general merchandise forms but did not have a general merchandise classification. All told, we have approximately 12,000 retailer-year observations (or about 2,400 per year), representing approximately 120,000 store-years (an average of about 10 stores per chain, or 24,000 stores per Census year). Of these, 10,000 store-year observations are single-units, and nearly 2,000 retailer-years are small chains, with 100 or fewer stores (averaging

[6]The Census send forms to all firms operating at least two establishments, even if the firm operates only a single retail store (other establishments operated by the firm may include manufacturing plants or wholesale establishments, for example). When we compute the size of the retailer, however, we limit the count to retail store operations.

[7]In some cases, part or all of the information may be provided by company headquarters. As noted below, due to non-response or incomplete responses not all stores that receive the forms are included in the data set.

Table 1. Chains' and Large Chains' Share of Sales

| | Chains' Share | | Large Chains' Share[c] | |
|---|---|---|---|---|
| Year | Retail[a] | GM[b] | Retail[a] | GM[b] |
| 1948 | 0.296 | 0.612 | 0.123 | n/a[d] |
| 1954 | 0.301 | 0.680 | 0.126 | 0.339 |
| 1958 | 0.337 | 0.748 | 0.143 | 0.360 |
| 1963 | 0.366 | 0.837 | 0.158 | 0.374 |
| 1967 | 0.398 | 0.867 | 0.186 | 0.461 |
| 1972 | 0.452 | 0.920 | 0.252 | 0.576 |
| 1977 | 0.480 | 0.945 | 0.268 | 0.622 |
| 1982 | 0.542 | 0.962 | 0.295 | 0.744 |
| 1987 | 0.546 | 0.969 | 0.309 | 0.757 |
| 1992 | 0.582 | 0.977 | 0.342 | 0.766 |
| 1997 | 0.600 | 0.981 | 0.372 | 0.921 |
| 2002 | 0.628 | 0.987 | 0.429 | 0.959 |

Source: Census of Business and Census of Retail Trade

[a] SIC 52 (to 1992) or NAICS 44-45 (from 1997)

[b] SIC 531 (to 1992) or NAICS 452 (from 1997)

[c] A large chain is a chain operating 100+ stores

[d] 1948 data on large chains' retail sales suppressed

Table 2. Retail Sector vs. General Merchandise Subsector, 2002

| | Retail[a] | GM[b] | GM/Retail |
|---|---|---|---|
| Sales (000,000,000$) | 2,975 | 433 | 0.146 |
| Retail Firms (000) | 727 | 9 | 0.013 |
| Stores (000) | 962 | 36 | 0.037 |
| Sales per Store (000$) | 3,092 | 12,038 | 3.893 |
| Stores per Retailer | 1.32 | 3.80 | 2.871 |
| Fraction of Stores in Chains | 0.38 | 0.78 | 2.061 |
| Fraction of Stores in Large Chains[c] | 0.23 | 0.72 | 3.101 |
| Sales Share of Top Four Retailers | 0.11 | 0.66 | 5.964 |

Source: Census of Business and Census of Retail Trade

[a] NAICS 44-45

[b] NAICS 452

[c] A large chain is a chain operating 100+ stores

8.5 stores per chain). Fewer than 200 chain-year combinations have more than 100 stores, and these average about 500 stores per chain.

Because the Census forms change from year to year, we created a concordance of product lines that is as consistent as possible over time. There are two types of product lines: "broad" lines and "detailed" lines, which provide a detailed breakdown of the broad lines. Table 3 lists the broad categories of goods, along with the years they are included on the forms and the number of detailed lines associated with each. The broad line "Groceries," for example, includes up to nine detailed lines: meat, fish, and poultry; fresh and prepackaged produce; frozen foods; dairy products; bakery products; deli items; soft drinks; candy; and all other foods. This level of detail is available only for 2002, however (and only for stores receiving form 45202); all other forms include the broad line "groceries" without the detailed breakdown.

The number and level of detail of lines listed on the Census forms has increased over time largely because the array of products carried by general merchandise stores has increased. Time series plots of lines carried can be misleading, however, because a line may have been carried by some, or many, stores for years before it first appeared on a Census form. To be conservative, we include year$\times$form fixed effects in all regressions, and exploit only the variation across retailer sizes to identify the effect of interest.

We think of a product line encompassing many substitutes, with minimal substitution *across* product lines. For example, there are many possible substitute outfits within the broad line Women's Apparel, but little substitution between Women's Apparel and Men's Apparel. Substitution across detailed lines within a broad line, such as Women's Suits and Women's Slacks, is more likely, but even here the definitions are, in general, broad enough so that a product's most common substitutes will lie within the same detailed product line.

With the concordance in place, counting the number of lines carried by any given store is straightforward, assuming that it reports lines data, although it is subject to reporting error. Stores vary in whether they submit a form at all (though it is required by law), whether they

Table 3. Broad Lines

| **Broad Line Description** | **Years** | **Details**[a] |
|---|---|---|
| Groceries | All | 9 |
| Meals, Snacks, and Nonalcoholic Beverages for Immediate Consumption | All | 1 |
| Packaged Liquor, Wine, and Beer | All | 3 |
| Tobacco Products and Accessories | All | 1 |
| Drugs, Health and Beauty Aids | All | 6 |
| Soaps, Detergents, and Household Cleaners | 1992-2002 | 1 |
| Paper and Related Products | 1992-2002 | 1 |
| Men's Apparel | All | 12 |
| Women's Apparel | All | 14 |
| Children's Apparel | 1992-2002 | 3 |
| Footwear | All | 5 |
| Curtains, Draperies, and Domestics | All | 2 |
| Major Household Appliances | All | 3 |
| Small Electrical Appliances | All | 1 |
| Televisions, VCRs, and Videotapes | All | 2 |
| Audio Equipment and Music | All | 3 |
| Furniture | All | 4 |
| Floor Coverings | All | 3 |
| Computer Hardware and Software | All | 2 |
| Kitchenware and Home Furnishings | All | 4 |
| Jewelry | All | 2 |
| Optical Goods (INcluding Eyeglasses, Telescopes, etc) | All | 1 |
| Sporting Goods (Including Bicycles and Guns) | All | 7 |
| Hardware, Tools, Plumbing, and Electrical Equipment and Accessories | All | 1 |
| Lawn and Garden Equipment and Supplies | All | 4 |
| Building Materials and Home Improvement Equipment and Supplies | All | 3 |
| Paint and Related Preservatives and Supplies | All | 1 |
| Automotive Supplies | All | 3 |
| Automotive Fuels | All | 1 |
| Household Fuels | All | 1 |
| Pets, Pet Foods, and Pet Supplies | 1997-2002 | 1 |
| Photographic Equipment and Supplies | All | 1 |
| Toys (Including Games and Crafts) | All | 2 |
| Sewing, Knitting, and Needlework Goods | All | 1 |
| Stationary, School, and Office Supplies | All | 2 |
| Luggage and Leather Goods | All | 1 |
| Office Equipment | 1987-2002 | 2 |
| Souvenirs and Novelty Items, Including Seasonal Decorations | 1997-2002 | 2 |
| Books, Magazines, and Newspapers | All | 2 |
| Miscellaneous Merchandise, Not Elsewhere Classified | All | 3 |
| Nonmerchanise Receipts | All | 8 |

[a] Maximum number of detailed lines associated with each broad line. Not all detailed lines are listed on each form each year, and some stores report only the broad line.

fill in any lines data (it is not uncommon to submit a form without that information), and the level of care and detail taken in filling out this information. Reporting error is therefore very likely, both in the count and identity of lines carried by each store, and in their revenue shares. Reporting error may be of two sorts: a store may report lines that it does not carry, or it may fail to report lines that it does carry. If the two types of error are equally likely, we have a noisy, but unbiased, measure of the number of product lines carried by each store. Stores that do not report any lines are not used in this part of the analysis. When a retail store reports selling a broad line but does not indicate how many detailed line(s) within the broad line it sells, we assign it 1 detailed line. (When we compute the number of stores in a chain, we use all stores, whether or not they have reported lines data.)

Counting the number of lines carried by each retail chain is trickier in the presence of reporting errors. When we compute the number of lines a chain carries, we use information only from the subset of stores in the chain for which the lines data are reported. Even so, it is common for some stores in a chain to report carrying a given product line — for example, exercise equipment — while others do not report it. The two leading explanations for this phenomenon are heterogeneity of product lines across stores, and reporting error. Because of the possibility of reporting error, the way we treat these partially-carried lines affects small and large chains differently. For example, counting only lines that are reported sold by 100% of stores in a chain would systematically undercount lines sold by large chains more than small chains or single-store retailers, because it only takes one store manager not filling out the form correctly to exclude the line from the data, and the larger the chain, the higher the probability of this sort of mistake. Conversely, counting all lines that are reported by *any* store within the chain, would lead to systematically over-counting the number of lines sold in large chains relative to small chains. (Consistent with this intuition, we find that the correlation between scale and scope is highest when we set the threshold for inclusion very low, and lowest when we set it high.) In the current analysis, we use three different thresholds — 25%, 50%, and 75% — for counting how many lines a retail chain carries.[8]

# 4 Empirical Findings

## 4.1 Correlations and Regressions

We start by documenting the relationship between number of stores in a chain and the number of lines the chain carries. We perform the analysis using store-level as well as chain-level data. Because we do not use sample weights in any of the analysis, the weights of various chain sizes depend on the level of the analysis: in the store-level analysis, all stores are treated equally, so the results are strongly influenced by chain stores, as they account for over 90% of the stores in the data; but in the retailer-level analysis, the relative share of chains shrinks to under 20%.

Chain stores, which increasingly dominate the general merchandise subsector (see Tables 1 and 2) tend to carry more distinct product lines than “mom and pop” general merchandise retailers; and the larger the chain, the more products do its individual stores carry.

The first row in Table 4 shows raw correlation coefficients between chain size and the number of broad lines the store (in the first column) or the chain (in the next three columns) carries. Table 5 shows the same correlations but uses detailed lines instead of broad lines. Across the full data set, at the store level, the correlation between the number of broad lines carried by the store and the number of stores in the chain to which the store belongs is 0.1524. The equivalent correlation for detailed lines is 0.0576. Counting product lines at the retail-chain level, we apply three different cutoff rules for the inclusion of each line — 25%,

[8] Retailer-level numbers also differ slightly due to our treatment of write-in and miscellaneous lines. When a store reports one or more lines that do not appear on the form, either as a “miscellaneous” line or by writing in the specific line’s description — e.g., pet food — we count this as both a single broad and a single detailed line at the store level. But because there is no way to compare write-in and miscellaneous lines across stores within a chain, we do not include those in the count of retailer-level lines regardless of the threshold used.

50%, and 75%; as expected, the correlation declines as the cutoff rule becomes stricter, but it is positive and statistically significant across the board.

Table 4. Broad Lines Results

| | | Chain-Level | | |
|---|---|---|---|---|
| Analysis | Store-Level | 25% Rule | 50% Rule | 75% Rule |
| Correlation | 0.1524*** | 0.1101*** | 0.1017*** | 0.0889*** |
| Regression | 0.0018*** | 0.0024*** | 0.0021*** | 0.0019*** |
| | (0.0000) | (0.0007) | (0.0007) | (0.0007) |
| Observations | 120,5713 | 11,985 | 11,985 | 11,985 |

Notes: Each cell represents a different regression. Asymptotic standard errors in parentheses. *** significant at 1%.

Table 5. Detailed Lines Results

| | | Chain-Level | | |
|---|---|---|---|---|
| Analysis | Store-Level | 25% Rule | 50% Rule | 75% Rule |
| Correlation | 0.0576*** | 0.1209*** | 0.1102*** | 0.0993*** |
| Regression | 0.0049*** | 0.0053*** | 0.0045*** | 0.0038*** |
| | (0.0001) | (0.0014) | (0.0014) | (0.0014) |
| Observations | 120,571 | 11,985 | 11,985 | 11,985 |

Notes: Each cell represents a different regression. Asymptotic standard errors in parentheses

Consistent with these correlations, single-store retailers carry, on average, 16.7 detailed lines or 12.0 broad lines. In contrast, stores belonging to chains with 2-99 stores carry, on average, 27.8 detailed lines or 16.7 broad lines, and stores belonging to chains with more than 100 stores carry 35.2 detailed lines or 21.1 broad lines.

To test whether chains tend to add product lines as they grow, we use store-level data to estimate

$$n_{ijft} = \sum_i \alpha_i + \sum_t \sum_f \delta_t \cdot \phi_{if} + \beta \cdot k_{jt} + \varepsilon_{ijft} \tag{1}$$

where $n_{ijft}$ is the number of (detailed or broad) lines of merchandise carried by store $i$ belong to retailer $j$ and receiving form $f$ in year $t$, $\alpha_i$ is a store fixed effect, $\delta_t$ is a year fixed effect (for 1982, 1987, 1992, 1997, and 2002), $\phi_{if}$ is a form fixed effect (for forms 5301, 5302, 45201, and 45202), and $k_{jt}$ is the number of stores operated by retailer $j$ in year $t$. The reason for including year×form fixed effects is that, as noted earlier, different forms contain different counts of detailed and broad lines, and these change over time. Estimates from this regression should not be interpreted causally, but instead should be interpreted as controlled correlations, i.e., correlations that allow us to control for the changing (generally, increasing) number of possible product lines a store manager can check off a form.

Estimates of the coefficient of interest, $\beta$, are shown in the second row of the first column in Tables 4 and 5 for broad and detailed line counts, respectively. Both coefficient estimates are significant at the 1% level. For detailed lines, the coefficient estimate is 0.0049, implying that, on average, whenever a chain adds one detailed line, it also adds 204 stores. The coefficient estimate for broad lines is 0.0018, so a chain that adds one broad line also adds, on average, about 555 stores.

We next estimate the effect at the chain level, using

$$n_{jt} = \sum_j \gamma_j + \sum_t \delta_t + \beta \cdot k_{jt} + \varepsilon_{jt} \tag{2}$$

with $\gamma_j$ now a retailer chain fixed effect. The subscript $f$ is removed because stores belonging to a single chain may receive different types of forms.[9] instead of a store fixed effects. These results are shown in the next three columns of the second row in both tables. The number of observations drops by an order of magnitude, increasing the standard errors in the regressions, but the coefficients are actually larger. Now, we estimate that (depending on the specification) the addition of 190-260 stores to a chain is associated with an increase of one detailed product line, and the addition of 430-530 stores is associated with an increase

[9]Because every year, both forms include all the same broad codes, the count of broad codes by retailer should not be affected by the heterogeneity of forms, but the level of detail does differ by form, making the detail-line analysis potentially sensitive to the distribution of forms a retail chain receives. In practice, however, most stores belonging to a single chain do receive the same form, so we do not think this is empirically important. In the regression equation, we could use the median form type or redefine $\phi_{jf}$ to be the *fraction* of stores operated by retailer to receive form $f$.

of one broad product line.[10]

Alternative specifications we plan to estimate in the future include a log-log specification, in which case we could interpret the coefficient $\beta$ as the (reduced-form) elasticity of lines with respect to chain size. Again, this coefficient could not be interpreted causally.

In the next section, we present a model in which $n$, the number of lines a store or chain carries, and $k$, the number of stores in the chain, are complements so that each is increasing in the other. If our model is correct, then the estimates of $\beta$ obtained so far are an upper bound on the *causal* effect of an increase in chain size on line counts.

# 5 A Model of Superstores

## 5.1 Environment

There are $L$ locations, which we think of as distinct, non-overlapping, markets (towns or neighborhoods). These locations are identical *ex ante*.

In each market $\ell$, $N$ normal goods are each supplied by one or more "mom-and-pop" (single product) retailers.[11] In a subset $k \in [0, L]$ of locations (with $k$ determined endogenously), there is also a chain "superstore" selling $1 < n \leq N$ goods, each at price $p$. The goods are perfectly symmetric with respect to demand and cost parameters, so the superstore sets a single price $p$ for all the goods it sells.

We do not explicitly model the consumer's decision problem and the competitive environment. Instead, we assume that in each location $k$ served by the chain, the inverse demand function $p(x, n)$ facing the chain is increasing in $n$, the number of products sold, and decreasing in $x$, the quantity sold of each product. The second assumption is standard. The first assumption reflects a "one stop shopping" (OSS) effect: the more goods the superstore sells,

[10] We have not completely cleaned up the data, and some issues remain with flagged observations that may need to be dropped. The regression results reported here appear to be robust to the inclusion or exclusion of these flagged observations.

[11] The assumption that mom and pop stores sell a single product is a normalization.

the more attractive is shopping there, and the higher the price the superstore can charge to sell a fixed amount of each good.[12] We also assume that the OSS effect declines as $x$ increases:

$$\frac{\partial^2 p}{\partial x \partial n} \geq 0.$$

Graphically, this assumption implies that the OSS effect not only shifts out the demand curve, but also makes it flatter (more elastic).

The chain superstore has three choice variables: the number of stores it operates (scale of the chain), the number of items it sells in every store (scope of the superstore), and the price it charges for each item (which determines the sales volume of the superstore). We assume that the number of stores in the chain, $k$, has a direct effect on its costs but not on consumer demand; the number of products, $n$, for sale in each store affects both costs and demand. The chain first determines its scale ($k$) and scope ($n$), and then sets its price.

To operate $k$ stores, the chain incurs a "chaining cost" $\Phi(k) \equiv \frac{\phi(k)}{\delta}$ where $\phi' > 0, \phi'' > 0$; $\delta$ is a technology parameter. In addition, the chain incurs a "scope cost" $Z(n)$, with $Z' > 0, Z'' > 0$, to sell $n$ unique goods. The motivation for the chaining cost is that the logistics and managerial problems associated with managing a chain become increasingly complex as the chain grows.[13] The scope cost has a similar motivation. We assume that the two costs are additively separable, except for a cost $R$ per product per store, which can be thought of as the rental rate per display aisle. The cost $R$ allows for some interdependence in the costs of $k$ and $n$, specifically, it allows for the cost of adjusting $k$ (respectively, $n$) to increase with the value of $n$ (respectively, $k$).

The superstore's profit function is

$$\pi = n \cdot (k \cdot x \cdot p(x, n; x_i) - C(kx)) - n \cdot k \cdot R - Z(n) - \frac{\phi(k)}{\delta} \tag{3}$$

[12]The notion that consumers have to pay some transportation cost or "distribution cost" to shop goes back to Hotelling (1929); Betancourt and Gautschi (1988) enumerate various types of distribution costs consumers may incur by shopping in multiple locations.

[13]See Basker and Van (2006) for a more detailed discussion.

where $p(x, n; x_i)$ is the inverse demand function for each product $n$ at each of the superstore's $k$ locations, $x$ is the quantity sold of each item, and $x_i$ is the quantity sold by fringe (mom and pop) competitors.

We assume that the cost of inputs $C(kx)$ is increasing, but at a decreasing rate, so that $C' > 0, C'' < 0$. There is ample evidence that retailers' average cost falls with volume. Although the Robinson-Patman Act (also known as the "Anti Chain-Store Act") has, since 1936, prohibited sellers from price discriminating where the effect may lessen competition, in practice large retailers pay lower prices per unit than smaller ones.[14] One common mechanism that generates cost differentials is manufacturers' practice of "reimbursing" large buyers for marketing expenses the retailers incur to promote their products. Our impression from conversations with retail industry insiders is that these payments depend more on the number of units a retailer sells than on any actual costs incurred; the per-unit "reimbursement" increases with the size of the retailer. In addition, if small and larger retailers purchase their wares from different sources, or sell differentiated items, the law may not apply. Chains with large sales volumes have additional potential for savings by contracting directly from overseas supplies, a choice that involves some fixed costs but lower marginal costs (Basker and Van, 2006); Gereffi (2006) presents some evidence that the largest apparel and general merchandise chains import a larger share of their apparel than do smaller retailers.

The timing of decisions is as follows. In the first stage, the chain choose its scale ($k$) and scope ($n$) and incurs the relevant costs. In the second stage, the chain and its fringe competitors simultaneously choose prices.[15] Consumers then observe prices and selection in all stores and make their shopping decisions.

We solve the problem by backwards induction, starting with the second stage. We start

[14]For a nice discussion about the Robinson-Patman Act's economic consequences, see Ross (1984).

[15]There are three reasons why price setting occurs after scale and scope decisions are made. First, it is more realistic, because scale and scope change only at low frequencies while prices can be changed frequently. Second, it will allow us later to model the competitive structure taking $n$ and $k$ as parameters. Third, is simplifies the algebra considerably.

with a simple case in which a perfectly competitive fringe does not price strategically. We then solve a more general problem in which the fringe competitors set prices strategically.

## 5.2 Price Setting with No Strategic Considerations

We assume here that, although there is only a single mom-and-pop store selling each product $i$, the market is contestable in that there is instantaneous free entry of an identical competitor, so each mom and pop store prices at marginal cost, $p_i = c$. In this case, the superstore's inverse demand function can be written as $p(x, n)$.

Define

$$m = x \cdot p(x, n) - \frac{C(kx)}{k}. \tag{4}$$

In the second stage, the superstore takes $n$ and $k$ as parameters, and maximizes $m$, average operating profit per store, with respect to $x$. The first-order condition implicitly defines $x^*(n, k)$:

$$x \cdot \left.\frac{\partial p}{\partial x}\right|_{x=x^*} + p(x^*, n) - C'(kx^*) = 0 \tag{5}$$

This quantity maximizes profit if the second-order condition

$$x \cdot \frac{\partial^2 p}{\partial x^2} + 2 \cdot \frac{\partial p}{\partial x} - k \cdot C''(kx) < 0 \tag{6}$$

holds everywhere. Let $m^*(n, k)$ be the maximized value of $m$; by the envelope theorem,

$$\begin{aligned}
\frac{\partial m^*}{\partial n} &= x^* \cdot \frac{\partial p}{\partial n} > 0 \\
\frac{\partial m^*}{\partial k} &= \frac{x}{k}\left(\frac{C(kx)}{kx} - C'(kx)\right) > 0.
\end{aligned}$$

Our first result states that the larger are the scope and scale of the superstore, the more units it sells of each good in equilibrium:

**Result 1** (Superstore Sells More Stuff). *Operating profit, $m$, is supermodular in $(x, n, k)$. It follows that $x^*(n, k)$ is monotone nondecreasing in $(n, k)$ and that $\frac{\partial^2 m^*}{\partial n \partial k} > 0$.*

*Proof.* Supermodularity requires that $m$ has increasing differences in $(x, n, k)$, or equivalently, since $m$ is continuous and twice differentiable, that $m$ has nonnegative cross-partial derivatives (Topkis, 1978).

$$\begin{aligned}
\frac{\partial^2 m}{\partial x \partial n} &= x \cdot \frac{\partial^2 p}{\partial x \partial n} + \frac{\partial p}{\partial n} \\
\frac{\partial^2 m}{\partial x \partial k} &= -xC''(kx) \\
\frac{\partial^2 m}{\partial n \partial k} &= 0
\end{aligned}$$

all of which are (weakly) positive by inspection.

Supermodularity of $m$ implies that $x^*(n, k)$ is monotone nondecreasing in $(n, k)$.

Finally, since $m$ has increasing differences in $(n, k)$ for all values of $x$, it must also have increasing differences in $(n, k)$ when evaluated at $x = x^*$. □

Figure 1 shows this result graphically.

Next, we relax the assumption that the competitive fringe has constant prices.

## 5.3 Price Setting in a Strategic Setting

If both the chain and the competitive "mom and pop" fringe set prices strategically, the demand function facing each store depends on its own prices as well as on prices at other stores. Equivalently, assuming the demand functions are invertible, the inverse demand function facing each store is a function of the quantity it sells as well as the quantity sold by its competitors.

For the results that follow, it is convenient to define the mom and pop store's choice variable as $y_i = -x_i$, where $y_i \in \mathbb{R}^-$ is the negative of the quantity sold by store $i$, and to rewrite the inverse demand functions accordingly. Let $p(x, n; y_i)$ be the superstore's inverse demand function, and let $p_i(y_i; x, n)$ be store $i$'s inverse demand function. We assume that $\frac{\partial p_i}{\partial x} < 0$ and $\frac{\partial p_i}{\partial n} < 0$, i.e., any increase in the number of products sold or number of units sold per product in the superstore shifts demand for the mom and pop store inwards. For tractability, we also assume that neither $p_i(y_i; x, n)$ nor $p_(x, n; y_i)$ depend on $x_j$, the quantity of any other good sold by another "mom and pop" store. Finally, we assume that any parameter that shifts the demand curve outwards also flattens it (making it more elastic), which implies the following restrictions on the cross-partial derivatives of demand:

$$\frac{\partial^2 p_i}{\partial n \partial y_i} \geq 0$$
$$\frac{\partial^2 p_i}{\partial x \partial y_i} \geq 0$$
$$\frac{\partial^2 p_i}{\partial x \partial n} \geq 0$$
$$\frac{\partial^2 p}{\partial n \partial y_i} \geq 0$$
$$\frac{\partial^2 p}{\partial x \partial y_i} \geq 0.$$

The chain as well as all the "mom and pop" stores take $(n, k)$ as parameters, and simultaneously maximize:

$$m(x; y_i, n, k) = x \cdot p(x, n; y_i) - \frac{C(kx)}{k} \tag{7}$$
$$\pi_i(y_i; x, n) = -y_i \cdot (p_i(x_i; x, n) - c) - \widetilde{R} \tag{8}$$

with respect to $x$ and $y_i$, respectively (where $x$ is nonnegative and $y_i$ is nonpositive); $c$ is the (constant) marginal cost of a mom and pop store, $\widetilde{R}$ is the rental (overhead) cost of the mom and pop store, and both $p$ and $p_i$ are increasing in $y_i$. Denote the superstore's best response (BR) function $x^*(y_i; n, k)$ and the mom and pop store's BR function $y_i^*(x; n, k)$.[16]

[16]In the competitive-fringe setting of the previous section, it is necessary to assume that $\widetilde{R} = 0$. This assumption is no longer necessary here, because the mom and pop store's price can be higher than marginal cost.

The next result states that $x$ and $y_i$ are strategic complements — alternatively, that $x$ and $x_i$ are strategic substitutes — and that the Nash equilibrium of this game, $(x^*(n,k), y_i^*(n,k))$, is increasing in $n$ and $k$: the larger is the superstore in terms of scale and scope, the more units does the superstore sell of each product, and the fewer its competitors sell in each store.

**Result 2** (Superstore Sells More Stuff and Shrinks Competitors)**.** *The above game is supermodular, implying that $x^*(n,k)$ and $y_i^*(n,k)$ are strategic complements. The chain's maximized average operating profit, $m^*$, is increasing in $(n,k)$. The game has smallest and largest pure Nash equilibria, denoted, respectively, $(\underline{x^*}, \underline{y_i^*})$ and $(\overline{x^*}, \overline{y_i^*})$, both of which are increasing in $(n,k)$.*

*Proof.* Supermodularity of the game requires that both $m$ and $\pi_i$ have increasing differences in $(x, y_i)$, or equivalently, since both functions are continuous and twice differentiable, that they have nonnegative cross-partial derivatives:

$$\begin{aligned}\frac{\partial^2 m}{\partial y_i \partial x} &= \frac{\partial p}{\partial y_i} + x \cdot \frac{\partial^2 p}{\partial y_i \partial x} \\ \frac{\partial^2 \pi_i}{\partial y_i \partial x} &= -y_i \cdot \frac{\partial^2 p_i}{\partial y_i \partial x}\end{aligned}$$

This implies that the game has at least one Nash equilibrium, and that the set of Nash equilibria has a smallest and largest element (Topkis, 1979).

In addition, $m$ is supermodular in $(x, n, k)$, and $\pi_i$ is supermodular in $(y_i, x, n, k)$:

$$
\begin{aligned}
\frac{\partial^2 m}{\partial x \partial n} &= x \cdot \frac{\partial^2 p}{\partial x \partial n} + \frac{\partial p}{\partial n} \\
\frac{\partial^2 m}{\partial x \partial k} &= -xC''(kx) \\
\frac{\partial^2 m}{\partial n \partial k} &= 0 \\
\frac{\partial^2 \pi_i}{\partial y_i \partial n} &= -y_i \cdot \frac{\partial^2 p_i}{\partial y_i \partial n} + \frac{\partial p_i}{\partial n} \\
\frac{\partial^2 \pi_i}{\partial y_i \partial k} &= 0 \\
\frac{\partial^2 \pi_i}{\partial n \partial k} &= 0
\end{aligned}
$$

all of which are (weakly) positive by inspection, given the above assumptions on the cross-partials of the inverse demand curve. This is sufficient to prove that the smallest and largest Nash equilibria are both increasing in $(n, k)$ (Milgrom and Roberts, 1990). Finally, since $m$ has increasing differences in $(n, k)$ for all values of $(x, y_i)$, it must also have increasing differences in $(n, k)$ when evaluated at an equilibrium $(x^*, y_i^*)$.

It follows that if the Nash equilibrium of this game is unique, it is increasing in $(n, k)$. □

This and the previous section therefore establish that the larger are the scale and scope of the retail chain, the more units it will sell of each good in each location, independently of the competitive structure of each market.

## 5.4 Chain Scale and Scope

In the first stage, anticipating these effects on second-stage sales, the superstore solves

$$
\max_{n,k} \quad \pi = kn \cdot (m^*(n, k) - R) - Z(n) - \frac{\phi(k)}{\delta}. \tag{9}
$$

where $m^*(n, k)$ is the maximized value of $m$ in the second stage of the game.

Our main result establishes that scale and scope are complements in the chain's profit function.

**Result 3** (Complementarity of Scale and Scope)**.** *On the domain where* $m^* \geq R$*,* $\pi$ *is supermodular in* $(n, k, \delta)$*. Therefore,* $(x^*, n^*, k^*)$ *are monotone nondecreasing in* $\delta$*.*

*Proof.* As above, we need show that the cross-partial derivatives of $\pi$ are nonnegative.

$$\begin{aligned}
\frac{\partial^2 \pi}{\partial n \partial \delta} &= 0 \\
\frac{\partial^2 \pi}{\partial k \partial \delta} &= \frac{\phi'(k)}{\delta^2} \\
\frac{\partial^2 \pi}{\partial k \partial n} &= (m^* - R) + n \cdot \frac{\partial m^*}{\partial n} + k \cdot \frac{\partial m^*}{\partial k} + n \cdot k \cdot \frac{\partial^2 m^*}{\partial n \partial k}
\end{aligned}$$

The first and second cross-partials are (weakly) positive by inspection. The third cross partial, $\frac{\partial^2 \pi}{\partial k \partial n}$, is the sum of four terms, the first of which is positive for $(n, k)$ such that $m^* > R$, and the second and third of which are unambiguously positive by Result 1 or 2. The last term is zero by either Result 1 or 2.[17] □

Figure 2 shows this result graphically.

This result implies that improvements in the retail chain's "chaining technology" — encompassing management, logistics and distribution technology — lead to larger retail stores.[18]

Analogously, any force that decreases the marginal cost — or increases the marginal benefit — of adding product lines to a store will increase the optimal store size. An example of a force that has decreased marginal costs is the introduction of bar codes (see Holmes,

[17] We may be able to do better. If we can show that

$$m^* + n \cdot \frac{\partial m^*}{\partial n} + k \cdot \frac{\partial m^*}{\partial k}$$

is bounded away from zero for all $n \geq 1$ and $k \geq 1$. In that case the result will be true for **R small enough** where "small enough" means smaller than the lower bound on the above expression.

[18] Similarly, trade liberalization, by increasing the economies of scale in purchasing, can lead to larger stores; see Basker and Van (2006) for a model relating trade liberalization and chain size.

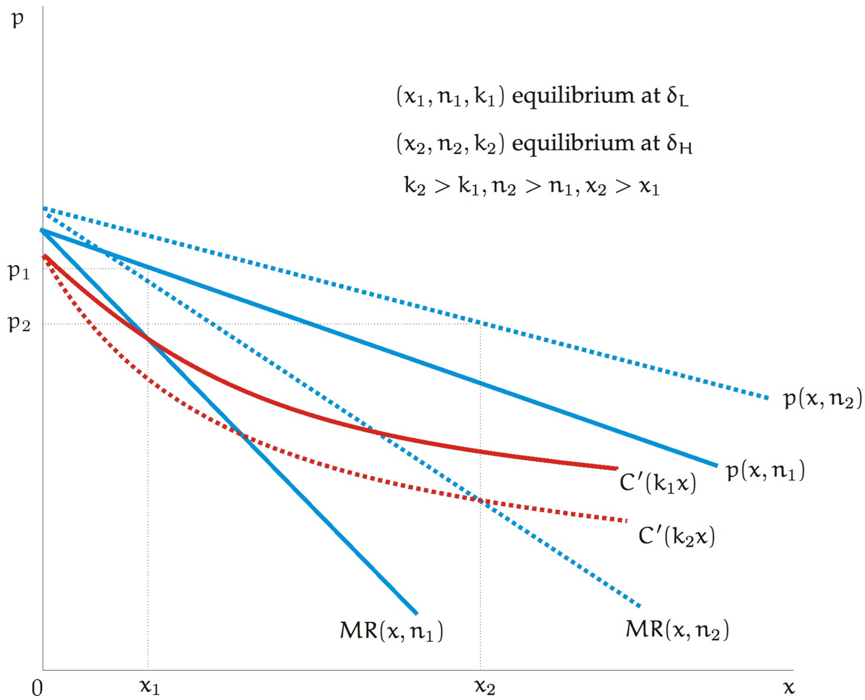


Figure 1. Second-Stage Equilibrium: $x^*(n, k)$

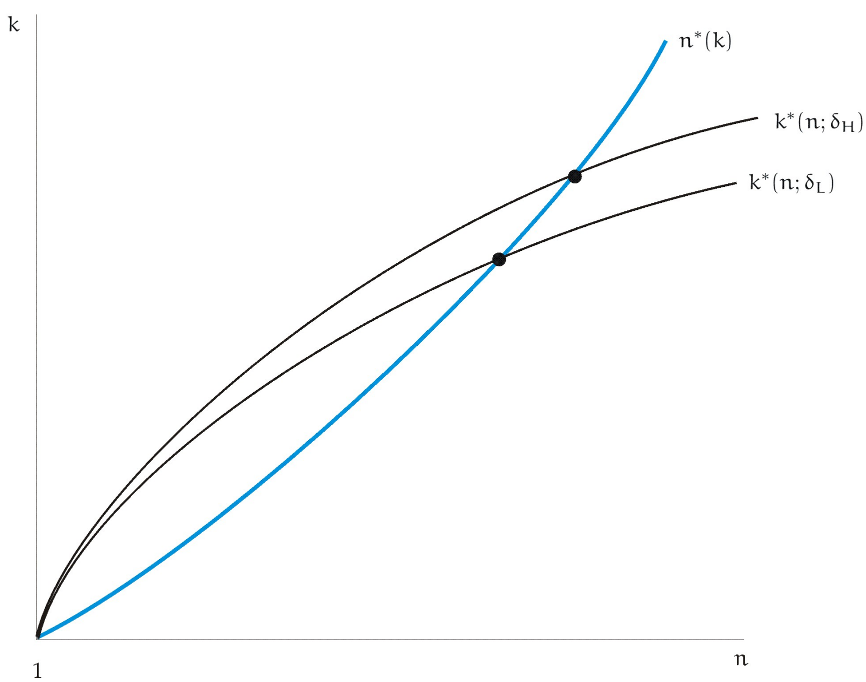


Figure 2. Global Equilibrium: $(n^*, k^*)$

2001). One example of a force that may have increased the marginal benefit of adding product lines is increased cost of time due to higher labor force participation of women; another example is suburbanization of demand, which has increased the optimal store size.[19]

# 6 Concluding Remarks

In this paper, we document, analyze, and offer an explanation for the simultaneous rise in chain size and product offerings by general merchandisers over the last several decades. The average number of distinct (detailed) product lines carried by stores in small chains (with 2-100 stores) is 50% higher than the average number of detailed product lines carried by single-store, or "mom and pop," retailers (28 vs. 17); it is 75% higher in stores that belong to large chains (with more than 100) than in single-store retailers. The figures for broad lines are qualitatively comparable, though the differences are somewhat smaller: stores in small chain carry 33% more broad lines, and stores in large chains carry about 50% more broad lines, than single-store retailers. In addition, the retail chains that have grown the most — adding the largest number of stores — over this period have also expanded into more product markets than other retailers. On average, retail chains have added one broad product line for every 500 stores over this period, or one detailed product line for every 200 stores.

Our model explains this relationship with an interaction of economies of scale and demand-driven gains from scope due to consumer preference for one-stop shopping. Diseconomies of both scale and scope prevent the retailer from exploding to fill all geographic markets and/or all product markets. The balance between the marginal cost and benefit to expanding product lines depends on the number of stores in the chain, because the larger the number of stores the larger the benefit of adding a product line. Similarly, the larger is

[19]See Pashigian and Bowen (1994) for other consequences of this trend for retailing

the number of product lines the retail chain carries, the greater is its profit in each store and therefore its benefit from expanding the number of stores.

Our model is a long-run equilibrium model, in that it does not account for short-term "stickiness" in the size of stores. In practice, store size cannot increase overnight, but if a retailer finds that its optimal store size has increased substantially, it has an incentive to close or remodel small stores. One way to further investigate these dynamics is to use another variable collected by the CRT, the square footage of stores. The extent to which increases in stores' product coverage and square footage tend to occur together, in discrete jumps, will give us an idea of the short-run (in)flexibility of retailers to adjust their product reach. We can also investigate the frequency of increases in product lines: at the store level, do lines increase steadily from one Census year to the next, or are there large jumps in a single year — possibly coinciding with a major renovation in which the store's square footage also rises — preceded, and followed, by relative stability?

While the count of lines increases, *ceteris paribus*, with a store's entry into a product line and decreases with exit, we do not explicitly address such entry and exit in the current draft. Addressing entry and exit of product lines will require a careful accounting of which product lines are included in which forms and in what years, so as to avoid misattributing a change in forms to a change in the store's product coverage.

Our paper has some interesting predictions and policy implications. In particular, our model implies that any exogenous force that increases store size — for example, consumers' increased preference for one-stop shopping due to changes in the composition of the labor-force or other changes in the value of time or the cost of gasoline — will lead chains to increase their scale as well as scope. Conversely, any exogenous force that increases a chain's optimal scale, such as improvement in technology, or cost reductions due to trade liberalization — will lead the chain to increase their scope as well as scale.

Complementarity between chain size and store size implies that local ordinances to limit store size could also limit the chain's size, and vice versa. In this light, some local ordinances that limit the expansion of **chain stores** — such as a measure passed by Chicago's City Council in 2006 (but later vetoed by the mayor) which would have doubled the effective minimum wage paid by stores with at least 90,000 square feet, operated by retailers with at least $1 billion in annual sales (Basker, 2006) — are better understood as attempts to combat large stores.

# References


Bagwell, K., G. Ramey, and D. F. Spulber (1997) "Dynamic Retail Price and Investment Competition," *RAND Journal of Economics*, 28(2), 207–227.

Barger, H. (1955) *Distribution's Place in the American Economy since 1869*. Princeton University Press, Princeton, NJ.

Basker, E. (2006) "The Causes and Consequences of Wal-Mart's Growth," unpublished paper, University of Missouri.

Basker, E., and P. H. Van (2006) "Putting a Smiley Face on the Dragon: Wal-Mart as Catalyst to U.S.-China Trade," unpublished paper, University of Missouri.

Betancourt, R., and D. Gautschi (1988) "The Economics of Retail Firms," *Managerial and Decision Economics*, 9(2), 133–144.

Bliss, C. (1988) "A Theory of Retail Pricing," *Journal of Industrial Economics*, 26(4), 375–391.

Foster, L., J. Haltiwanger, and C. J. Krizan (2006) "Market Selection, Reallocation and Restructuring in the U.S. Retail Trade Sector in the 1990s," *Review of Economics and Statistics*, 88(4), 748–758.

George, L., and J. Waldfogel (2003) "Who Affects Whom in Daily Newspaper Markets?," *Journal of Political Economy*, 111(4), 765–784.

Gereffi, G. (2006) "The New Offshoring of Jobs and Global Development," unpublished manuscript, Duke University.

Holmes, T. (2001) "Bar Codes Lead to Frequent Deliveries and Superstores," *RAND Journal of Economics*, 32(4), 708–725.

——— (2005) "The Diffusion of Wal-Mart and Economies of Density," unpublished paper, University of Minnesota.

Hotelling, H. (1929) "Stability in Competition," *Economic Journal*, 39(153), 41–57.

Jarmin, R. S., S. D. Klimek, and J. Miranda (2005) "The Role of Retail Chains: National, Regional and Industry Results," unpublished paper, U.S. Census Bureau.

Loertscher, S., and G. Muehlheusser (2006) "Global and Local Players in a Model of Competition," unpublished paper, University of Bern.

Milgrom, P., and J. Roberts (1990) "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58(6), 1255–1277.

Pashigian, B. P., and B. Bowen (1994) "The Rising Cost of Time of Females, the Growth of National Brands and the Supply of Retail Services," *Economic Inquiry*, 32(1), 33–65.

Ross, T. W. (1984) "Winners and Losers under the Robinson-Patman Act," *Journal of Law and Economics*, 27(2), 243–271.

Topkis, D. M. (1978) "Minimizing a Supermodular Function on a Lattice," *Operations Research*, 26(2), 305–321.

——— (1979) "Equilibrium Points in Nonzero-Sum $n$-Person Submodular Games," *SIAM Journal on Control and Optimization*, 17(6), 773–787.